What is claimed is:

1. An utterance detector comprising:

a frame-level detector for making speech/non-speech decisions for each frame, and

an utterance detector coupled to said frame-level detector and responsive to said speech/non-speech decisions over a period of frames to detect an utterance.

2. The utterance detector of Claim 1, wherein said utterance level detector is a state machine.

3. The utterance detector of Claim 2, wherein said state machine has the states of pre-speech, non-speech, in-speech and pre-non-speech.

4. The utterance detector of Claim 1, wherein said frame-level detector includes autocorrelation.

5. The utterance detector of Claim 4, including filter means for performing frequency-selective autocorrelation.

6. The utterance detector of Claim 5, wherein said autocorrelation and filtering is performed in DFT domain by taking the signal and applying DFT, performing frequency domain windowing and then inverse DFT.

7. The utterance detector of Claim 1, wherein said frame-level frame detector includes means for calculating power spectrum of an input signal, performing frequency shaping, performing inverse FFT and determining maximum value of periodicity.

8. The utterance detector of Claim 7, wherein calculating power spectrum includes the steps of filtering the signal, applying a Hamming window and performing FFT on the signal from the Hamming window.

9. The utterance detector of Claim 7, wherein said performing frequency shaping step includes the step of

$$F(k) = \begin{cases} \alpha^{F_l - k} & \text{if} \quad 0 \le k < F_l \\ 1 & \text{if} \quad F_l \le k < F_h \\ \beta^{k - F_h} & \text{if} \quad F_h \le k < \dfrac{N}{2} \end{cases}$$

with $\alpha = 0.70$

$\beta = 0.85$

to get R(k).